LA-UR-18-29909
Approved for public release; distribution is unlimited.

Title: Verification and Validation of Optimization of Replacement Portal Monitoring Replay Tools

Author(s): Klasky, Marc Louis
Perry, John Oliver
McKerns, Michael

Intended for: Report

Issued: 2018-10-18

Verification and Validation of Optimization of Replacement Portal Monitoring Replay Tools

The Los Alamos National Laboratory and the Johns Hopkins Applied Physics Laboratory (APL) are both developing optimization software to fulfill the objective of Phase III of the Replacement Portal Monitoring (RPM) Program for the Department of Homeland Security. The objective of the Phase III of the RPM Program is to conduct an evaluation of three RPM systems designed to enhance mission effectiveness, gain operational efficiencies, and to address emerging mission needs against the Countering Weapons of Mass Destruction (CWMD) nuclear threat basis and special nuclear material (SNM) used in performance testing. As such this task involves the use of global optimization algorithms.

Global optimization (GO) methods can roughly be classified as deterministic and stochastic strategies. Stochastic methods for global optimization ultimately rely on probabilistic approaches. Given that random elements are involved, these methods only have weak theoretical guarantees of convergence to the global solution. However, these methods generally can locate the vicinity of the global solution with relative ease.

Deterministic methods are those that can provide a level of assurance that the global optimum will be located. However, it should be noted that, although deterministic methods can guarantee global optimality for certain GO problems, no algorithm can solve general GO problems with certainty in finite time.
[1]

Because the complexity of the manifold of the solution space for the RPM Portal Replay Tools is unknown, general statements regarding the obtainment of the global minima cannot in general be obtained. Accordingly, we employ multiple optimization strategies to examine the minima. Furthermore, the ability to reduce the dimensionality of the data set by utilizing clustering techniques enables us the opportunity to explore the solution space more completely and consequently increases the likelihood of achieving the global minima. However, it must be cautioned that the clustering of the data increases the coarseness of the solution space and hence many more solutions may be possibly obtained with equivalent minima than one in which no clustering is performed.

To ensure that the algorithms developed by the respective labs are able to determine the global minima several test problems are proposed. The first test problem involves using all of the Characterization Test (CT) data for each of the respective vendor systems. This data set includes items which are characterized as items to be detected as well as items are that are benign items i.e. items if detected would indicate a false alarm. This test problem avoids the generation of synthetic data and other data processing steps. The metric to be evaluated is the number of items detected versus undetected. This evaluation will be at 4 levels of false alarm. The function to be minimized in the evaluation will have the functional form[2]:

$$F\left(P_{d,}P_{fatarget,}\right) = (1 - m)(1 - P_d) + m \text{ ABS } ((P_{fa} - P_{fatarget}) / P_{fatarget}) \quad (1)$$

Where:

$$m(P_{fa}, P_{fatarget}) = 1 - \exp(- (P_{fa} - P_{fatarget})\text{^}2 / (2 \text{*} sigma\text{^}2))$$

σ~0.1*Pfatarget.

Another test problem to evaluate the global minimization algorithms will involve the use of a synthetic data set generated by APL. Details of this data set will be provided by APL. Briefly, APL will supply LANL with the spectra for which the minimization of the cost function will be evaluated. Again, four different settings for the probability of false alarm will be evaluated.

An addition test problem using the clustered set of the Stream of Commerce (SoC) data in conjunction with the APL synthetic SNM spectra will also be examined. The clustered set of SoC will be developed by LANL and provided to APL for this evaluation. APL will provide the set of source synthetic spectra. The cost function will again be evaluated at four prescribed false-alarm probabilities.

Finally, we address the issue of comparing the solution sets i.e. the parameter settings of the vendors Replay Tools obtained during the minimization of the cost function in the multi-objective optimization problem.

$$Minimize\ (f_1(x), f_2(x), f_M(x)) \qquad \text{Equation 1}$$
$$subject\ to\ x \epsilon\ S$$

It is envisioned that many solutions with the same global minima i.e. cost function will be obtained based upon preliminary evaluations by LANL of the solution space of the CT data. However, it remains to be determined if the different optimization strategies will uncover solution sets with the same parameter settings. In this vein a question that may be raised is that of the preferred solution. Simply stated, these solutions are non-dominated by one another. That is, there exist no other solution in the entire search space that dominate any of these solutions. Thus, if the parameters are perfectly implemented into the black-box algorithms each of the solutions is the optimal solution. On the other hand if the parameters cannot be perfectly implemented then the resulting global optima solution may be effected. Consequently, if a global solution is quite sensitive to such variable perturbations in its vicinity, the implemented solution may result in a different set of objective values than that of the theoretical optimal solution. Thus, we seek solutions that are referred to as robust and that do not exhibit or minimize this property. Much work has been performed in finding such robust solutions.[3], [4],[5] Essentially instead of optimizing the original function we optimize the mean effective function computed by averaging a representative set of neighboring solutions. Solutions which are less sensitive to such local perturbations will perform well in terms of the mean effective objective values and the resulting efficient front may turn out to be the robust frontier.

Mathematically we define a robust solution for a single objective function as:

$$Min\ f^{eff}(x) = \frac{1}{[\beta_\delta(x)]} \int_{y \epsilon \beta_{\delta(x)}} f(y)dy \quad Equation\ 2$$
$$subject\ to\ x\ \epsilon\ S$$

where:

β_δ *is the* δ *neighborhood of the solution* x *and* $[\beta_\delta(x)]$ *is the hypervolume neighborhood*

Implementation of Equation 2 requires a finite set of solutions H may be either be randomly or structurally sampled, i.e. Latin hypercube method, chosen around a δ neighboordhod B $\delta(\mathbf{x})$ of a solution x in the variable space and the mean function value (f_{eff}). Since this cause H times as many evaluations the use of a Linux based operating system which will allow for on the order of 10-100 greater processing power is recommended. Additionally, the use of a dynamically updated archive of a fixed size for choosing neighboring solutions is recommended and offers faster computation.[6] As suggested by Deb, the approach utilized is to calculate the normalized difference in values between the perturbed function value f_p (can be f_{eff}) and the original function f itself and declare a solution to be robust, if the normalized difference is smaller than a chosen threshold (η).[7]

Extension of the concept of the single robust optimization to multi-objective functions is performed by examining each Pareto-optimal solution to its insensitivity towards small perturbations in its decision variable values. The extension of the single objective robust solution is performed by recognizing:

1. The sensitivity is established with respect to all M objectives. That is, a combined effect of variations in all M objectives has to be used as a measure of sensitivity to variable perturbation.
2. There are many solutions to be check for robustness to multi-objective optimization as opposed to one or two for single-objective optimization.

$$Min\ (f_1^{eff}(x), f_1^{eff}(x), f_1^{eff}(x), f_1^{eff}(x), Equation\ 3$$
$$subject\ to\ x\ \epsilon\ S,$$

Where:

f_j^{eff}(x) is defined as follows:

$$f_j^{eff}(x) = \frac{1}{[\beta_\delta(x)]} \int_{y \epsilon \beta_\delta} f_j(y)dy$$
$$subject\ to\ x\ \epsilon\ S$$

Due to the variable sensitivities, a part of the original global Pareto-optimal (efficient) front may not qualify as a robust front. In some scenarios, the original global efficient front (corresponding to the problem stated in Equation 3) may be completely non-robust and an original local efficient or an original sub-optimal front may now become robust. Depending on how robust the original global efficient front is with respect to the above definition, there can be the following four main scenarios[8]:

• **Case 1:** The complete original efficient front is robust.
• **Case 2:** A part of the original efficient front is no more robust.
• **Case 3:** The complete original global efficient front is non-robust; instead an original local efficient front is robust.
• **Case 4:** A part of the original global efficient front is robust together with a part of an original local efficient front.

We illustrate and discuss each of the above four scenarios in the following. Later, we develop one test problem for each scenario.

Case 1

This is the simplest case in which the original efficient front remains as an efficient front with respect to the mean effective objective functions. Figure 3 illustrates such a problem for a two-objective optimization problem. Although it is expected that the global efficient front constructed with the mean effective objectives will be somewhat worse than that constructed with the original objectives, the entire set of original Pareto- optimal solutions is robust and is the target in this type of optimization problems. The concept can also be extended for more than two objectives.

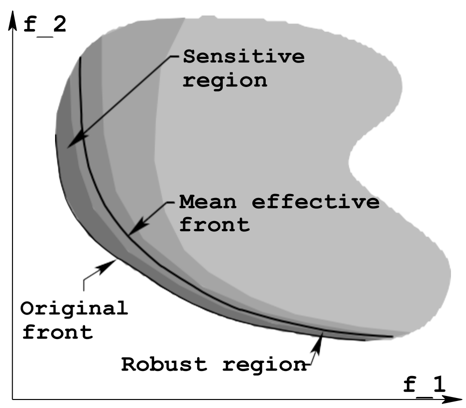

Figure 3: Case 1: Complete efficient front is robust.

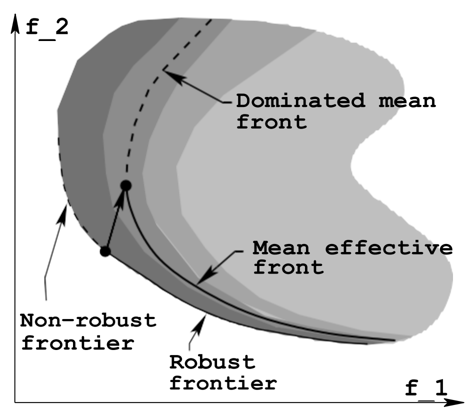

Figure 4: Case 2: A part of the efficient front is r

Case 2

Here, the entire original efficient front is not robust with respect to the above definition for robustness of type I. In most real-world scenarios such a problem is expected, as some portion of the original efficient front may lie in a sensitive region in the decision variable space.

In such a problem, an important task of a robust optimizer would be to identify only that part of the efficient front which is robust (that is, less sensitive to the variable perturbation). Figure 4 shows that the efficient front corresponding to the mean effective objectives does not span over the entire original efficient region in a two-objective optimization problem.

Case 3

Cases 3 and 4 correspond to more difficult problems in which the original problem may have more than one efficient frontiers (global and local). In a Case 3 problem, the global efficient front of the original problem is completely dominated by a local efficient front with respect to the mean effective objectives, thereby meaning that the original global Pareto-optimal solutions are not robust solutions and that they are sensitive to local perturbation. Figure 5 demonstrates such a problem with two objectives. This type of problem, if encountered, must be solved for finding the robust efficient front, instead of the sensitive global efficient front.

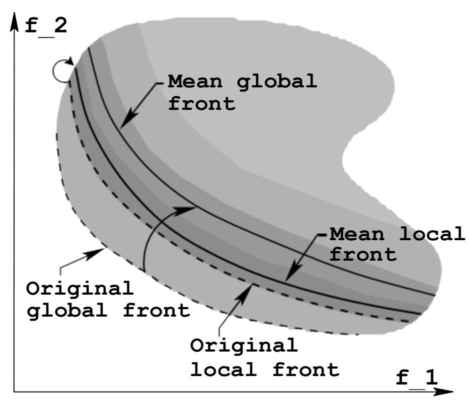

Figure 5: Case 3: The complete global efficient front is not robust.

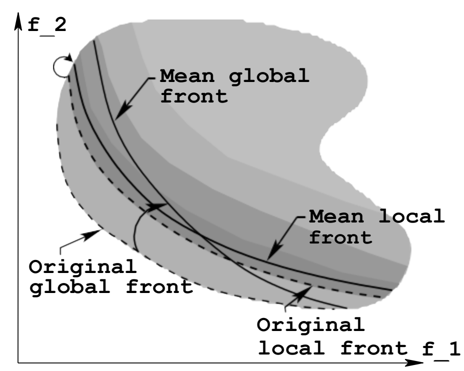

Figure 6: Case 4: A part of the global efficient front is not robust.

Case 4

Instead of the complete original global efficient front being sensitive to the variable perturbation, Case 4 problems cause a part of it to be adequately insensitive. In the remaining part, a new front appears to be robust. Figure 6 illustrates this problem with two objectives. A part of the robust frontier corresponds to the original global efficient frontier and the rest corresponds to the original local efficient frontier.

Certainly, other scenarios are possible, where instead of an original local efficient front becoming robust, a completely new frontier emerges to be robust. This is likely in the case of constrained optimization with Pareto-optimal solutions lying on the constraint boundary. The concepts can also be extended to problems having more than two objectives.

While the previous description involves the examination of the minima to perturbations in the Replay tool parameters it is not envisioned that they cannot be implemented exactly. Therefore, the applicability of the robust optimization solution to the RPM Optimization is not clear. However, the problem that is apparent in determining the global minima is that of the evolving data set. That is, although the Replay Tool parameters are determined from one sample of data within the entirety of possible data the data presented at any given period may differ from the data set utilized to determine the original parameters. To address this issue we propose to segment the data from each of the respective sites into X separate batches and compare the parameter settings for each respective batch. We can also investigate the sensitivity of the parameters to different segments of the clustered SoC data. A general formalism to address this problem can be developed but is outside of the scope of this Phase of the project.

---

[1] Rehman, S., Langelaar, M., "Efficient global robust optimization of unconstrained problems affected by parametric uncertainties, Struc Multidisc Optim (2015) 52:319-336
[2] D.Portnoy et al., NIM A 652 pp29-32, 2011
[3] *Ben-Tal, A.; Arkadi Nemirovski, A. (2002). "Robust optimization—methodology and applications". Mathematical Programming, Series B.* ***92****: 453–480.*
[4] *Chen, X.; Sim, M.; Sun, P. (2007). "A Robust Optimization Perspective on Stochastic Programming". Operations Research.* ***55*** *(6): 1058–1071.*
[5] Kouvelis P. and Yu G. (1997). *Robust Discrete Optimization and Its Applications,* Kluwer.
[6] Branke, J., and Schmidt, C., "Faster convergence by means of fitness estimation", Parallel Problem Solving from Nature, pages 119-128, 1998.
[7] Deb, K. and Gupta, H. "Introducing Robustness in Multi-Objective Optimization", KanGAL report
[8] Ibid 7